Exhibit 6

Continued

Exhibit C

DA SILVA MOORE ET AL. V. PUBLICIS GROUPE SA ET AL. CASE NO. 11-CV-1279

PLAINTIFFS as of 11/14/12

NAME	TITLE/LOCATION/ DATES OF EMPLOYMENT	DATE OF CONSENT TO JOIN
Amanda Rosen	Senior Vice President New York June 2004 – April 2011	09/12/12
Becky Lauer	Senior Vice President New York November 2000 – April 2011	07/12/12
Carol Perlman	Senior Vice President New York January 2007 – June 2011	08/18/11
Diana Scott	Vice President New York February 2001 – April 2010	09/20/12
Ellyn Caravetta	Senior Vice President New York May 2000 – August 2011	09/10/12
Emily Buchanan	Senior Vice President New York September 2007 – August 2011	10/15/12
Erica Bersin	Senior Vice President New York October 2010 – March 2011	10/08/12
Erin Libit	Vice President Washington D.C. April 2008 – June 2011	10/08/12
Erin Ortiz	Vice President New York July 2000 – October 2012	10/15/12
Heather Pierce	Vice President Washington D.C. / San Francisco October 2004 - December 2008	n/a
Holly Richmond	Vice President Atlanta October 2004 – July 2010	09/10/12
Joellen Zumberge	Senior Vice President London May 1996 – January 2010	09/20/12

DA SILVA MOORE ET AL. V. PUBLICIS GROUPE SA ET AL. CASE NO. 11-CV-1279

PLAINTIFFS as of 11/14/12

Kate Wilkinson	Account Executive Los Angeles February 2008 - March 2010	n/a
Kelly McKenna	Vice President San Francisco May 2003 – June 2011	08/24/12
Krista Webster	Senior Vice President New York January 2000 – December 2009	09/27/12
Larissa Severenko	Vice President New York February 2003 – March 2010	10/14/12
Laura Ann Hall	Senior Vice President New York September 2010 - July 2011	10/11/12
Laurie Mayers	Senior Vice President Ann Arbor March 2002 - April 2010	n/a
Lori Laurent Smith	Senior Vice President Ann Arbor February 2010 – February 2011	09/27/12
Margy Meislin	Senior Vice President New York October 1988 – March 1992 June 2007 - July 2010	09/25/12
Maryanne Caruso	Senior Vice President New York April 2004 – January 2011	09/21/12
MaryEllen O'Donohue	Senior Vice President New York June 1985 - February 2010	n/a
Megan Jordan	Senior Vice President Los Angeles March 1999 – January 2003 February 2007 – August 2010	10/14/12
Melanie Babcock	Senior Vice President Atlanta March 2008 – June 2012	10/01/12

DA SILVA MOORE ET AL. V. PUBLICIS GROUPE SA ET AL. CASE NO. 11-CV-1279

PLAINTIFFS as of 11/14/12

Michelle Overall	Vice President New York January 2008 – May 2010	10/13/12
Monique da Silva Moore	Senior Vice President Boston April 1991 - March 1993 June 2004 - January 2010	n/a
Nancy Brenner	Senior Vice President New York January 1996 - Current	10/17/12
Nyree Wright	Vice President Washington D.C. April 2011 – Current	10/15/12
Robyn Leventhal	Vice President New York May 2003 – April 2011	09/20/12
Sheila McLean	Senior Vice President Washington D.C. March 2002 – Current	10/13/12
Stephanie Andrzejewski	Senior Vice President New York March 2000 – August 2004 September 2005 –December 2010	09/10/12
Wendy R Clark	Senior Vice President Ann Arbor June 2010 – August 2012	09/10/12
Zaneta Hubbard	Account Supervisor Atlanta February 2008 - December 2008	n/a

Exhibit D

From: Siham Nurhussein
To: Publicis
Subject: FW: da Silva Moore
Date: Wednesday, November 14, 2012 2:31:33 PM

From: Anders, Brett M. (Morristown) [mailto:AndersB@JacksonLewis.com]
Sent: Wednesday, November 14, 2012 2:30 PM
To: Siham Nurhussein
Cc: Susan Rubenstein; Janette Wipper; Brecher, Jeffrey W. (Long Island); Chavey, Victoria Woodin (Hartford)
Subject: RE: da Silva Moore

Siham,

Please see our responses to your proposal below. As you will see, there are a few areas where we disagree as well as a few areas where we need clarification from you. To that end, please provide clarification as soon as possible.

In addition, and as previously discussed, the following is a list of the individuals we understand supervised the opt-in Plaintiffs and whose e-mail accounts we intend to collect. Please advise as soon as possible if you have any proposed modifications or corrections to this list.

1. Stephanie Smith
2. Jeanine O'Kane
3. Matt Gardner
4. Joe Carberry
5. Bill Orr
6. Anne DeSchweinitz
7. Kelly Dencker
8. Katie Adams
9. Joel Curran
10. Ellen F. Schneidau
11. Wendy Lund
12. Margy Meislin
13. Stephanie Koze
14. Michael Morsman
15. Tom Vickery
16. Chuck Alston
17. Renee Wilson
18. Rob Baskin
19. Kyle Farnham
20. Catherine Falcetti
21. Elise Titan
22. Don Hannaford
23. Jim Tsokanos

24. Neil Dhillon
25. Bruce Mackenzie
26. Vicki Fite
27. Kelly Walsh
28. Christine Abbott
29. David Bashaw
30. Anita Bose
31. Caryn Carmer Previdi
32. Dan McGinn

Please get back to me as soon as possible since we need to respond to Judge Peck by tomorrow, November 15, 2012.

Brett M. Anders
Attorney at Law
Jackson Lewis LLP
220 Headquarters Plaza
East Tower, 7th Floor
Morristown, NJ 07960

973-451-6305 | Direct
973-538-6890 | Main
973-540-9015 | Fax

andersb@jacksonlewis.com

www.jacksonlewis.com

From: Siham Nurhussein [mailto:SNurhussein@sanfordheisler.com]
Sent: Tuesday, November 13, 2012 5:36 PM
To: Brecher, Jeffrey W. (Long Island); Susan Rubenstein
Cc: Anders, Brett M. (Morristown); Chavey, Victoria Woodin (Hartford); Janette Wipper
Subject: RE: da Silva Moore

Hi Jeff,

Following up on our call yesterday, we are providing you with our proposed changes to the ESI protocol. As we noted yesterday, we believe the current protocol needs to be tweaked in light of the addition of 26 new opt-in plaintiffs. Since the existing protocol has proven to be unexpectedly time-consuming as well as burdensome to the parties and the Court, we also propose streamlining the process via a much simpler protocol. Our proposal, set out below, will reduce the number of joint steps from 9 to 2 while shortening the time schedule:

a. MSL will update its database to include, e.g.:

i. The addition of all opt-in plaintiffs as custodians (26 additional custodians). **MSL Response: We will agree to the inclusion of the opt-in Plaintiffs as custodians in the database of documents to be reviewed. However, as set forth in our October 29, 2012 letter to Judge Carter, we do not believe the following individuals should be included as opt-in Plaintiffs, and thus, custodians: Ellyn Caravetta, Maryanne Caruso, Margy Meislin, Lori Laurent Smith, Erica Bersin and Laura Hall. Therefore, we will await the Court's guidance as to whether these individuals are proper opt-in Plaintiffs and, correspondingly, whether they should be included as custodians for purposes of e-discovery.**

ii. Broader temporal scope to capture discoverable information from the new opt-in plaintiffs, some of whom are current employees. Consistent with FRCP 26(e), MSL should supplement its ESI production beyond its proposed end-date for the collection of e-mail accounts (November 1, 2012). **MSL Response: We do not agree to further extending the collection period beyond November 1, 2012.**

b. MSL will produce a new random sample for joint review, which will "provide the parties with an updated baseline from which to gauge the effectiveness of the predictive coding process." Letter, B.M. Anders to D. Bains, November 5, 2012. MSL will also produce a new set of seed set documents for joint review. To minimize the burden on the Court, any disputes regarding the coding of the jointly reviewed documents (initial random sample and seed set) will be resolved by a Special Master, to be agreed upon by the parties. **MSL Response: We will agree to producing a new initial random sample. Please confirm that you propose a Confidence Level of 95% with a Confidence Interval of +/- 5% for the size of the random sample. (See comment below.) As it relates to a new seed set, we have been advised that the preparation of a new seed set is not necessary. The current seed set comprises of approximately 16,000 documents which we understand is sufficient for the initial training of the software. As we conduct the iterative training rounds, we will use the various tools within Axcelerate to further train the software as to the inclusion of the opt-in Plaintiffs. However, in light of your statement that the opt-in Plaintiffs have varying forms of Title VII claims (e.g., failure to promote, pregnancy, etc.), for each opt-in Plaintiff, please provide us with a summary of the nature of their Title VII claims. This information is necessary to enable us determine the types of documents which may be relevant to their alleged claims. We will need this information prior to starting the review of the new initial random sample. Lastly, we do not agree to the use of a Special Master. Practically speaking, Judge Peck has already ruled on a significant number of relevancy issues which were brought to the Court's attention through the development of the seed set. As a result, the key areas of dispute have been ruled on already and, based on his intimate knowledge of the case, any further disputes as it relates to relevancy can be resolved relatively quickly.**

a. MSL will create a complete production set using whatever methods and procedures it deems appropriate, including, should it so desire, consulting Plaintiffs as to relevancy, etc. **MSL Response: This is acceptable to the extent that it**

contemplates that MSL will be responsible for conducting the iterative rounds of training. However, given the parties' disagreements regarding proportionality and the appropriate number of documents to be reviewed/produced when weighing the benefit versus the burden and expense, a final production set cannot be completed until after the parties agree, or obtain a ruling from the Court, concerning the appropriate number of documents to be manually reviewed for final production. As you may recall, Judge Peck clearly stated that he would apply the principles of proportionality to this case and that he would make a threshold determination as to the number of documents that MSL had to review and produce. Therefore, MSL proposes that after it completes the iterative rounds of predictive coding, it makes a proposal as to the number of documents to be reviewed for production (e.g., the top ranked 40,000 documents – or whatever number agreed to by the parties or determined by the court). MSL will provide an explanation as to the basis for its proposal. At that point, but prior to final review and production, MSL will generate and the parties will review a second random sample comprised of the remaining documents (e.g., everything other than the top-ranked 40,000 documents – or whatever number agreed to by the parties or determined by the court) to determine whether it contains any relevant documents and, if so, the nature of those documents (e.g., are they highly relevant or "more of the same.")

b. Upon completion of the production set, MSL will generate
    i. a simple random sample from the entire collection, produced and unproduced, of a size sufficient to ensure that it will with high probability contain sufficient responsive documents to allow estimating a 95% confidence interval on recall with a margin of error of 0.05 on a recall scale of 0.0 to 1.0; **MSL Response: As set forth above, we believe this random sample should be conducted prior to the final review of documents, especially since the parties may need guidance from the Court regarding appropriate number of documents to be reviewed. Also, our position is that this random sample should be drawn from the "null set" or those documents we do not intend to manually review as part of the final review process. Lastly, please confirm that, as it relates to the size of the random sample, what you mean to say is that we will use a sample size that will allow a standard statistical Confidence Level of 95% and a Confidence Interval of +/-5%.**
    ii. a list specifying the unique ID of every document in the simple random sample and whether or not it is included in the production set; **MSL Response: This is moot in light of the comment above that the random sample be drawn from the null set.**
    iii. a list specifying the unique ID of every document in the collection and the score or scores assigned to it by the final predictive model or models used by MSL. **MSL Response: We will need to get back to you regarding this proposal. In theory, it is acceptable, however, we may need to modify the language.**
c. Representatives of the parties will make a responsiveness determination on each

document in the simple random sample. Documents in the simple random sample should be reviewed in a random order, and neither the production status nor the score(s) assigned by predictive modeling should displayed to personnel making responsiveness assessments. Any disputes regarding the coding of the final random sample will be resolved by a Special Master. **MSL Response: Generally, this is acceptable. However, for the reasons mentioned above, we do not agree to the use of a Special Master. Also, the production of all documents within the random sample to Plaintiffs' counsel is with the understanding that any privileged documents or any highly sensitive non-responsive documents may be withheld.**

d. Each party will make whatever statistical calculations it deems appropriate based upon the responsiveness determinations and production set inclusion status, including recall, precision, elusion, etc. **MSL Response: This is acceptable.**

e. Based on the estimates of production set effectiveness from the random sample, a determination will be made as to whether further searching for responsive documents is necessary. Any disputes regarding this issue will be resolved by the Court. **MSL Response: As set forth above, MSL's position is that this should be done prior to the final manual review of documents.**

f. If and when additional documents are added to the collection and evaluated for production, the procedure will be repeated, treating the new documents as a new collection to be produced from and evaluated by a new sample. **MSL Response: Can you explain further what you mean by this statement? For example, are you stating that if additional e-mail accounts are added in the future, the process above will be repeated? If so, that process may be overly cumbersome depending on the volume of additional documents. In the event that additional e-mail collections will need to be reviewed, the parties should meet and confer regarding an appropriate review method based on the number of documents at issue.**

With the abovementioned tweaks to the protocol, the parties should be able to complete the e-discovery process much more quickly. We propose the following timetable:

30 Days: Parties review new random sample and seed set documents
30 Days: MSL to conduct iterative rounds of predictive coding **MSL Response: This is acceptable.**
60 Days: Parties review final random sample to validate predictive coding **MSL Response: This should be reduced to 30 days, and then an additional 30 days should be allocated to the final review of the top-ranked documents (the number of documents to either be agreed to by the parties or ordered by the court).**
5 Days: Automated privilege review and production. **MSL Response. We do not agree to an automated privilege review. As a result, this period should be increased to 15 days.**

Please let us know as soon as possible whether you agree with the proposal outlined above. If there are particular aspects with which you disagree, please indicate so and provide your reasons.

Thanks,

Siham

From: Brecher, Jeffrey W. (Long Island) [mailto:BrecherJ@jacksonlewis.com]
Sent: Monday, November 12, 2012 5:09 PM
To: Susan Rubenstein; Siham Nurhussein
Cc: Anders, Brett M. (Morristown); Chavey, Victoria Woodin (Hartford)
Subject: da Silva Moore

Susan—

Confirming our conversation today, you will speak with your expert regarding our November 5, 2012 letter and provide us with changes to our proposal or the ESI protocol suggested by your expert. Because we must submit a letter to Judge Peck on November 15, please get back to us as soon as possible so we have an opportunity to review with our expert. Also, you agreed to provide us with a list of supervisors for the opt-in Plaintiffs so we can discuss collecting their e-mail. Thanks.

Jeffrey W. Brecher
Jackson Lewis LLP
58 South Service Road, Suite 410
Melville, NY 11747.
Phone: (631) 247-4652
Fax: (631) 247-0417 or 0418
brecherj@jacksonlewis.com

Brett M. Anders
Attorney at Law
Jackson Lewis LLP
220 Headquarters Plaza
East Tower, 7th Floor
Morristown, NJ 07960

973-451-6305 | Direct
973-538-6890 | Main
973-540-9015 | Fax

andersb@jacksonlewis.com

www.jacksonlewis.com
Representing management exclusively in workplace law and related litigation.

Exhibit E

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

MONIQUE DA SILVA MOORE, MARYELLEN O'DONOHUE, LAURIE MAYERS, HEATHER PIERCE, KATHERINE WILKINSON on behalf of themselves and all others similarly situated, and ZANETA HUBBARD, on her own behalf.	)	
PLAINTIFFS,	)	Civ No. 11-CV-1279 (ALC) (AJP)
v.	)	
PUBLICIS GROUPE SA and MSLGROUP,	)	
DEFENDANTS.	)	

DECLARATION OF DAVID D. LEWIS IN SUPPORT OF PLAINTIFFS' PROPOSED CHANGES TO THE ESI PROTOCOL

David D. Lewis states as follows:

1. I am a citizen and resident of the United States of America. The facts stated in this declaration are within my own personal knowledge and, if called as a witness to testify, I could and would competently testify to the facts contained in this declaration.

2. I am a consulting computer scientist and President of David D. Lewis Consulting, LLC, in Chicago, IL. I have been retained by Sanford Heisler, LLP as a consulting expert in the above-captioned matter.

3. I hold a Ph.D. From the University of Massachusetts at Amherst in Computer Science. My Ph.D. dissertation on using supervised learning for text classification (the core technology of "predictive coding") won the 1992 American Society for Information Science Doctoral Forum Award.

4. Between 1992 and 2000 I held research positions at the University of Chicago, Bell Labs, and AT&T Labs. Since 2000, I have been an independent consultant, and have consulted for more than 40 clients on problems in text search, text classification, statistical evaluation, natural language processing, data mining, and related areas.

5. I have published over 75 articles, including over 30 peer-reviewed conference and journal articles, on topics in information retrieval and statistical evaluation. I am an inventor on 8 issued patents in these areas.

6. I was elected a Fellow of the American Association for the Advancement of Science in 2006. The citation on my election read "*You are being honored for contributions to machine learning for automated indexing, including algorithms, evaluation frameworks, and datasets which are used worldwide by computer and information scientists.*"

7. In 2006, I was the co-founder (with Doug Oard and Jason Baron) of the TREC Legal Track, the first open evaluation of search and classification technology for e-discovery. Prior to that, I was the designer of the first TREC text classification evaluations, the TREC-4 and TREC-5 Filtering Tracks. More generally, I have served on the organizing committees of more than 15 United States government evaluations of information retrieval and natural language processing technologies.

8. I submit this declaration to point out serious flaws in the strategy proposed by Defendants for evaluating the effectiveness of predictive coding, as well as issues with their use of predictive coding. I also discuss how the evaluation strategy proposed by Plaintiffs will allow unbiased estimation of the effectiveness of both predictive coding and the review process as a whole, including the estimation of effectiveness measures of interest put forward by both Plaintiffs and Defendants.

9. I am a proponent of predictive coding, and have spoken widely to both legal and computer science audiences in support of this technology and its ability to reduce costs and produce better results in e-discovery. I am equally a proponent, based on over 20 years of experience with text classification technology, for careful and unbiased statistical evaluation of the effectiveness of any application of text classification. The recent increases in the number of plaintiffs in this case, and the corresponding increase in the size and diversity of documents to be searched, make a proper statistical evaluation of effectiveness even more important.

Defendants' Proposed Sampling Strategy Cannot Distinguish Between Effective and Ineffective Use of Predictive Coding

10. Defendants propose that predictive coding be applied to the collected documents in this case, and used to identify a subset of the collection (which we will call the "prioritized set") to be manually reviewed. The remainder of the collection would not be reviewed unless found by happenstance (e.g. by the accident of belonging to a random sample). Therefore responsive documents from this portion of the collection, which we will refer to (using Defendants terminology) as the "null set", would not be produced unless found by happenstance.

11. Defendants propose (as recently as the 14-Nov-2012 email from Mr. Anders to Plaintiffs' attorneys, attached to Plaintiffs' November 15, 2012 letter as Exhibit D) that

the effectiveness of predictive coding can be demonstrated by drawing a simple random sample from the null set, and comparing the results with a previously drawn simple random sample from the entire collection. This can not be relied upon, as is easily shown by an example.

12. An initial simple random sample of 2,399 documents, reduced to 2,388 documents after privilege screening, has already been jointly reviewed by Plaintiffs and Defendants. This sample was drawn from a collection of documents that at the time held approximately 3.2 million documents. The joint assessments of this sample resulted at different points in time in anywhere from 17 to 49 responsive documents. For the purpose of my analysis, I will assume the mean of these two values, 33 responsive documents, exist in the sample.

13. The sample prevalence (proportion of responsive documents in the sample) was therefore 0.0138, and this would be the usual point estimate of the collection prevalence. A 95% confidence interval on the collection prevalence based on this sample is [.0091, .0185]. (Unless otherwise stated, numerical results in this declaration are from the calculator at *http://www.raosoft.com/samplesize.html*. This calculator uses a gaussian approximation to the hypergeometric distribution. The aforementioned confidence interval is based on confidence level 95%, population size 3,200,000, response distribution 33/2388 = 1.38%, and sample size 2388.)

14. Defendants have proposed reviewing no more than 40,000 documents, so I will consider scenarios where the prioritized set contains 40,000 documents. I assume the collection contains 3,200,000 documents, but the analysis is not sensitive to moderate variations in collection size. I consider two possible scenarios.

15. In Scenario 1, the actual (and unknown) division of the collection imposed by predictive coding has 30,444 true positives, 9,556 false positives, 28,756 false negatives, and 3,131,244 true negatives. Thus in Scenario 1 the actual (and unknown) collection prevalence is (30,444+28,756)/3,200,000 = 0.0185, the upper end of our confidence interval on prevalence. The actual (and unknown) recall of predictive coding, i.e. the proportion of responsive documents included in the prioritized set , is 30,444/(30,444+28,765) = 0.5143. Whether this would be considered a successful use of predictive coding can be debated. But certainly predictive coding in Scenario 1 has had some success at concentrating responsive documents in the 40,000 to be reviewed.

16. In Scenario 2, the actual (and again unknown) division of the collection imposed by predictive coding has 364 true positives, 39,636 false positives, 28,756 false negatives, and 3,131,244 true negatives. In this scenario, the actual collection prevalence is (364+28,756)/3,200,000 = 0.0091, the lower end of our confidence interval on prevalence. In Scenario 2, predictive coding is completely useless: the proportions of responsive documents in the prioritized set, 364/40,000 = 0.0091, and in the null set, 28,756/(28,756+3,131,244) = 0.0091, are both the same as in the collection as a whole. The predictive coding system essentially chose documents at random. Recall in this scenario is 364/(364+28,756) = 0.0125.

17. However, in both Scenario 1 and Scenario 2, the number of documents in the null set (3,131,244), and the number of responsive documents in the null set (28,756), is exactly the same. Further, the collection prevalence in both Scenario 1 and Scenario 2 is consistent with the 95% confidence interval estimate for collection prevalence. Thus the two Scenarios cannot be distinguished by Defendants' sampling strategy. ***The uncertainty in the original prevalence estimate means that no approach based on sampling only from the null set can be relied upon to distinguish good from poor predictive coding.***

Defendants' Proposed Test for Whether a Review Has Achieved High Effectiveness is Invalid

18. In addition to the fundamental flaw with Defendants' sampling approach, the particular use they propose making of the resulting samples, while not described clearly, itself appears flawed.

19. Mr. Saggebruch's Declaration of March 7, 2012 states: "*The efficacy of the review will be shown through statistical sampling: an initial set of 2,399 documents will be reviewed to determine the expected responsiveness rate. By reviewing this number of documents, there will be a 95% confidence level (+/- 2%) in the expected responsiveness rate.* ***After the iteration cycles are complete, a second set of 2,399 will be reviewed to determine the decrease in responsiveness rates. A higher decrease in the responsiveness rate equates to a more effective review***" [my emphasis].

20. What appears to be proposed here is a comparison of the prevalence of the responsive documents in the collection with the *elusion* (the prevalence of the responsive documents in the null set). Defendants do not commit to a concrete effectiveness measure, but imply their measure would be related to the decrease in prevalence between the original collection and the null set.

21. Without a precise definition of how Defendants' intend to formalize a decrease in prevalence as an effectiveness measure, I cannot analyze it in detail. Whatever the measure is, however, its value could only be estimated from samples, not known exactly, unless the entire collection were to be jointly assessed. This makes the statistical properties of the measure, whatever it is, important.

22. I can make one straightforward observation about any measure Defendants could define based on decrease in prevalence. Consider again a scenario where Defendants' predictive coding approach simply behaves randomly, such that elusion (prevalence in the null set) is identical to collection prevalence. Assume that an initial simple random sample is drawn from the collection and a later simple random sample is drawn from the null set. Since the prevalence of responsive documents in the null set is the same as the prevalence of responsive documents in the collection, there is a 50% probability that prevalence within the sample drawn from the null set will be lower than the prevalence

within the sample drawn from the collection. **Thus 50% of the time in this situation where predictive coding was completely useless, Defendants' approach would indicate that there was some evidence of the success of predictive coding.**

23. A simple comparison of sample prevalence values therefore cannot serve as a test for the success of predictive coding. Defendants at a minimum would need to define a precise measure of decrease, determine its sampling distribution, and propose how to compute confidence intervals (or hypothesis tests) for it.

Other Flaws in Defendants' Approach to Evaluation

24. Defendants propose comparing prevalence values from two samples that are assessed at different points in time, potentially by different personnel. Further, assessors would be aware that in the first case the documents are from a representative sample of the collection, while in the second case that documents were rejected by a predictive coding system asserted by Defendants to have high effectiveness. Both of these contextual factors potentially bias the manual assessments in a way that would make prevalence rates computed from the two samples incomparable. Given the apparently low prevalence in the collection, a change in the assessment of even a small number of documents could have a major impact on the computation of any effectiveness measure of interest.

25. Another problem with Defendants' approach to evaluation is that sampling from the null set for a particular prioritization produced by predictive coding only allows evaluation of that single predictive model. Of even more interest is evaluating the effectiveness of Defendants' entire production process, including all methods of finding responsive documents, and any losses incurred through the error of reviewers (if Defendants choose to use manual review in their production process). Defendants' approach to evaluation does not allow this latter effectiveness to be estimated.

Obvious Variants of Defendants' Approach to Evaluation Are Also Flawed

26. There are several variants one can imagine of Defendants' quality control approach that do not address the basic flaw, or which introduce new problems.

27. First, in addition to a simple random sample from the null set, a separate simple random sample of the prioritized set could also be assessed. The two samples could be randomly ordered and assessed by assessors who do not know which of the samples each document came from. Taken together, the two samples would constitute a stratified random sample from the entire collection.

28. A stratified sample can be used to estimate recall (as proposed by Plaintiffs), some measure of the change in prevalence (as proposed by Defendants), or any other effectiveness measure for binary classification. However, estimating confidence intervals

from a stratified sample may require complex statistical procedures and software. A recently accepted journal article (Webber, William. Approximate Recall Confidence Intervals. *http://arxiv.org/abs/1202.2880*) spends forty pages discussing techniques for estimating recall from stratified samples, and finds that the best methods are those using randomized simulation techniques. Estimating Defendants' effectiveness measure, once they define it, would require work to understand that measure, and develop stratified estimation techniques for it, a potentially nontrivial basic research project in statistical science. Thus, while stratified sampling is a valid approach, and potentially could decrease sample size, there would be implications for the cost and risk of failure of the entire evaluation.

29. A second variant on Defendants' proposal would be to do joint assessment of *all* documents in the prioritized set, as well as a simple random sample of the null set. This would simplify estimation. However, even under Defendants' assumptions, this would require reviewing on the order of 40,000 documents, and Plaintiffs may rightfully argue that this set should be much larger. Thus this would be an expensive proposition.

30. A third variant would be to simply assume the correctness of Defendants' assessment of responsiveness for documents in the prioritized set, and use joint assessments of a sample from the null set only. This would produce hopelessly biased estimates of any effectiveness measure: it *assumes* what quality control should be attempting to *measure*.

Plaintiffs' Proposed Sampling Strategy Allows Estimating Effectiveness Measures Proposed by Both Defendants and Plaintiffs

31. Plaintiffs propose using the standard strategy for evaluating classification systems: take a simple random sample of the entire collection large enough to contain a reasonable number of responsive documents, jointly assess it for responsiveness by assessors who are blinded as to which documents have been prioritized, and estimate effectiveness measures based on this sample. This approach to evaluation is widely used in other applications of text classification outside e-discovery, and more broadly for applications of other types of data classification in industry and science.

32. A simple random sample of the entire collection can be used to estimate recall, precision, or any other effectiveness measure for binary classification. While Defendants have not defined what effectiveness measure they propose to use, it appears that elusion is a component of their approach. Plaintiffs' proposed sampling strategy would allow elusion to be estimated essentially as well as Defendants' own strategy, while allowing the estimation of other effectiveness measures that Defendants' strategy would make impossible to ever know.

33. For instance, consider Defendants' proposal to limit review to 40,000 documents, and to take a simple random sample of 2,399 documents from the null set. If the sample proportion were 0.01, the worst case margin of error for a confidence interval on elusion based on Defendants' sampling strategy would be 0.0040 (*raosoft* arguments: population

size 3,200,000-40,000 = 3,160,000, response distribution 1%, sample size 2,399). If instead the simple random sample of 2,399 were drawn from the entire collection, on average 2,369 documents of that sample would fall in the null set. Those 2,369 documents would constitute a simple random sample from the null set, and so could be used for estimating elusion. Again assuming a sample proportion of 1%, the margin of error on an estimate of elusion from a sample of 2,369 documents would still be 0.0040, identical to four decimal places (*raosoft* arguments: population size 3200000-40000 = 3160000, response distribution 1%, sample size 2369).

34. Thus under their own assumptions, Defendants' sampling strategy would not provide a meaningful advantage over Plaintiffs' sampling strategy, even in estimating an effectiveness measure *put forward by Defendants*. What Defendants' proposed strategy *would* do is to make it impossible to estimate recall, i.e. make it impossible to determine how effective predictive coding has been at prioritizing responsive documents for review.

Plaintiffs' Evaluation Strategy Can Be Implemented with A Reasonably Sized Sample

35. Plaintiffs propose that a simple random sample be drawn from the collection, of sufficient size to allow estimation of recall (the proportion of all responsive documents that are concentrated in a designated subset of the collection) with 95% confidence and a margin of error of 0.05 (5%). Plaintiffs' primary interest is estimating recall and other effectiveness measures for the production process as a whole, including the impact of any errors in manual review. However, the same sample could be used for estimating recall and other effectiveness measures for, say, a particular predictive model.

36. To guarantee 95% confidence and a margin of error of 0.05 on recall in the most difficult case (recall of 0.50) requires that the simple random sample be large enough to contain 385 responsive documents. To guarantee 95% confidence and a margin of error of 0.10 would require the sample to contain 97 responsive documents.

37. The initial random sample discussed in paragraph 12 contained between 17 and 49 responsive documents, depending on when assessments were made. We again assume 33 responsive documents out of 2,388 documents total. This gives a 95% confidence interval on the collection prevalence of [.0091, .0185]. Taking the lower end of this range suggests a quality control sample size of roughly 385/0.0091 =~ 42,308 documents. A small excess could be added to this to reduce the probability of fewer than 385 responsive occurring in the sample. For a margin of error of 0.10, this approach suggests a quality control sample size of roughly 97/0.0091 =~ 10,659 documents.

38. However, several techniques are available to reduce the size of the quality control random sample. One is to the select the documents sequentially, stopping when, say, 385 responsive documents have been found. This provides considerable reduction in sample size if the actual prevalence is toward the high end of the estimated range. The sequential

approach has the disadvantage that it slightly biases the estimates of effectiveness measures other than recall, though the bias is small and corrections are possible in some cases.

39. Another strategy would be to use stratified sampling, as discussed in paragraph 28. This introduces complexities in estimation as discussed, but would reduce sample size to some degree.

40. The final, and most effective, strategy for reducing the size of the quality control sample is for Defendants to actually achieve a higher level of recall. Suppose the recall of Defendant's production is 0.80 or higher. Then a margin of error of 0.05 can be achieved with a sample containing only 246 responsive documents rather than 385, leading to an average sample size under lower end prevalence of 246/0.0091 =~ 27,033. For a margin of error of 0.10, the corresponding figures would be 62 responsive and 62/0.0091 =~ 6,813 average sample size. Achieving higher levels of recall would allow even smaller sizes for the quality control sample.

Defendants' Proposed Predictive Coding Strategy Can Not Be Relied Upon to Achieve Good Effectiveness

41. Unfortunately, Defendants' proposed predictive coding strategy can not be relied upon to achieve high recall, or for that matter to achieve any particular level of effectiveness by any measure.

42. Defendants have consistently argued for arbitrary limitations on the number of documents to be reviewed. Disturbingly, they propose delaying the selection of a quality control sample until *after* an arbitrary limit on review has been established. Mr. Anders 14-Nov-2012 email states: '*Therefore, MSL proposes that after it completes the iterative rounds of predictive coding, it makes a proposal as to the number of documents to be reviewed for production (e.g., the top ranked 40,000 documents – or whatever number agreed to by the parties or determined by the court). MSL will provide an explanation as to the basis for its proposal. At that point, but prior to final review and production, MSL will generate and the parties will review a second random sample comprised of the remaining documents (e.g., everything other than the top-ranked 40,000 documents – or whatever number agreed to by the parties or determined by the court) to determine whether it contains any relevant documents and, if so, the nature of those documents (e.g., are they highly relevant or "more of the same.")*'

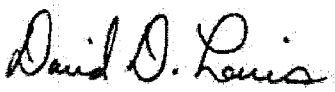
43. This is truly putting the cart before the horse: the best guidance to choosing how many documents to review is a simple random sample from the entire collection of documents, i.e. a quality control random sample selected in the fashion proposed by Plaintiffs. Plaintiffs have also requested that they be provided with a list specifying the unique ID of every document in the collection and the score or scores assigned to it by the final predictive model or models used by Defendants. This will provide additional guidance as to the number of documents to review and, in combination with the quality

control random sample, allow statistically validating any claim that the effectiveness of predictive coding has reached any purported "cliff edge". I concur with Judge Peck's advice during the Feb. 8, 2012 hearing: "*I'm not saying that there is a dollar number that I'm going to cut it off at or a percentage or where the cliff is. We are going to figure all that out.*" Plaintiffs' proposal provides the best means to figure all that out.

44. It is also unclear if the training strategies proposed by Defendants are designed to achieve good effectiveness. Mr. Anders 14-Nov-2012 email states "*As it relates to a new seed set, we have been advised that the preparation of a new seed set is not necessary. The current seed set comprises of approximately 16,000 documents which we understand is sufficient for the initial training of the software.*" However, these 16,000 documents were selected from what is now only a subset of a much larger collection. Further, the responsiveness assessments made on those documents are no longer valid. While there was discussion during the 14-May-2012 hearing about how the use of category judgments could be used to compensate for the change in responsive definition from, for instance, the addition of additional plaintiffs, the details of this are not clear to me or to Plaintiffs.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed Nov. 15, 2012, at Chicago, Illinois



David D. Lewis

Exhibit F

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

MONIQUE DA SILVA MOORE, MARYELLEN O'DONOHUE, LAURIE MAYERS, HEATHER PIERCE, KATHERINE WILKINSON on behalf of themselves and all others similarly situated, and ZANETA HUBBARD, on her own behalf.	
PLAINTIFFS,	**Civ No. 11-CV-1279 (ALC) (AJP)**
v.	
PUBLICIS GROUPE SA and MSLGROUP,	
DEFENDANTS.	

DECLARATION OF DOUGLAS E. FORREST IN SUPPORT OF PLAINTIFFS' PROPOSED CHANGES TO THE ESI PROTOCOL

I, Douglas E. Forrest, declare:

1. I am the Director, Discovery Analytics & Management at International Litigation Services ("ILS"), which is located in Aliso Viejo, California (www.ilsteam.com). I have been retained as a consultant for the Plaintiffs in this action. The facts stated in this declaration are within my own personal knowledge and, if called as a witness to testify, I could and would competently testify to the facts contained in this declaration.

2. I was admitted to the bar in 1977, and, after practicing law at Breed, Abbott & Morgan and Cravath, Swaine & Moore, where I worked extensively on discovery matters, I developed significant expertise in computer technology design and implementation, both generally and with respect to litigation support and e-discovery.

3. As an attorney at Cravath, I relied on Aquarius, the first large-scale implementation of computerized litigation support, which was implemented on the IBM antitrust cases.

4. As Director of Litigation Services at Legal Information Technology, Inc., I was instrumental in introducing imaging and coding for discovery to law firms, and pioneered in integrating imaging with legacy systems such as *BRS*.

5. As a systems architect, application designer and programmer, I created case management, litigation support and document repository systems (including *WIDE*, and *LIT CaseWorks for Lotus Notes*), SaaS (Software as a Service) knowledge management applications (including *LexisNexis Total Alerts* and *LexisNexis Clipper*) and e-Discovery and production operation systems. As the Chief Technology Officer of Ozmosys, I designed Knowledge Management (KM) based systems, including *Alerts Manager*, *Clipper* and *Open Alerts*. I also was the eDiscovery Services Manager for the Professional Services Division of Guidance Software, the creators of *EnCase*, the best known and most widely used forensic and enterprise eDiscovery software. I have been involved in, designed and managed numerous discovery and e-discovery projects.

6. At ILS, I direct discovery analytics and have directly participated in predictive coding, the application of other advanced analytic techniques to discovery materials and advising in the drafting of predictive coding protocols.

7. I submit this declaration in support of Plaintiffs' proposed changes to the ESI protocol.

A Revised Protocol Will Reduce Costs and Time

8. The Court's 10/11/2012 Order directed that "Counsel also should discuss any appropriate changes to the e-discovery protocol or procedures related to the opt-ins."

9. As set out in the Declaration of Deepika Bains, submitted herewith, the existing protocol has proven to be unexpectedly time consuming as well as burdensome to the parties and the Court itself, and, as discussed in the Declaration of David D. Lewis, also submitted herewith, is ill suited to handle the almost 500% increase in the number of plaintiffs as well as further increases from class certification or additional plaintiffs in the absence of such certification. (Similar issues are likely to arise from any changes required due to differences in the characteristics of documents included in Phase I vs. those included in Phase II.)

10. Accordingly, Plaintiffs propose to streamline the process with a substantially simpler protocol which would:

a. reduce the number of time-intensive joint review stages from 9 to 2,

b. relieve the burden faced by the Plaintiffs in proceeding under a protocol which is so deeply intertwined with a particular proprietary technology with which they have no first-hand experience, and

c. ensure that the final random sample data provides a basis to calculate all the statistics and metrics required to validate the results of the production or desired by either party.

11. Defendants have agreed to revise the protocol, however disagree as to some proposed aspects.

Plaintiffs' Proposed Protocol

12. Plaintiffs propose the following revisions:

1. MSL will update its database to include, e.g.:
    a. The addition of all opt-in plaintiffs as custodians (26 additional custodians), and
    b. Broader temporal scope to capture discoverable information from the new opt-in plaintiffs, some of whom are current employees. Consistent with FRCP 26(e), MSL should supplement its ESI production beyond its proposed end-date (November 1, 2012) for the collection of e-mail accounts.
2. MSL will produce a new random sample for joint review, which will "provide the parties with an updated baseline from which to gauge the effectiveness of the predictive coding process." (Letter, B.M. Anders to D. Bains, November 5, 2012.) MSL will also produce a new set of seed set documents for joint review. To minimize the burden on the Court, any disputes regarding the coding of the jointly reviewed documents (initial random sample and seed set) will be resolved by a Special Master, to be agreed upon by the parties.

3. MSL will create a complete production set using whatever methods and procedures it deems appropriate, including, should it so desire, consulting Plaintiffs as to relevancy, etc.
4. Upon completion of the production set, MSL will generate
    a. a simple random sample from the entire collection, produced and unproduced, of a size sufficient to ensure that it will with high probability contain sufficient responsive documents to allow estimating a 95% confidence interval on recall with a margin of error of 0.05 on a recall scale of 0.0 to 1.0;
    b. a list specifying the unique ID of every document in the simple random sample and whether or not it is included in the production set; and
    c. a list specifying the unique ID of every document in the collection and the score or scores assigned to it by the final predictive model or models used by MSL.
    d. Representatives of the parties will make a responsiveness determination on each document in the simple random sample. Documents in the simple random sample should be reviewed in a random order, and neither the production status nor the score(s) assigned by predictive modeling should be displayed to personnel making responsiveness assessments. Any disputes regarding the coding of the final random sample will be resolved by a Special Master.
5. Each party will make whatever statistical calculations it deems appropriate based upon the responsiveness determinations and production set inclusion status, including recall, precision, elusion, etc.
6. Based on the estimates of production set effectiveness from the random

sample, a determination will be made as to whether further searching for responsive documents is necessary. Any disputes regarding this issue will be resolved by the Court.

7. If and when additional documents are added to the collection and evaluated for production, the procedure will be repeated, treating the new documents as a new collection to be produced from and evaluated by a new sample.

13. The most time and cost saving element of the proposal is eliminating the seven 500 document review iterations to be performed jointly. The most critically important element is ensuring that the final random sample provides a basis upon which all relevant statistics can be generated and the quality of the entire process thereby be validated.

14. Because Plaintiffs' proposed protocol would not require the parties to jointly participate in the iterative review, Plaintiffs' willingness to play a smaller role in this phase of the protocol should be contingent on the adoption of appropriate quality control standards.

Eliminating the Jointly Reviewed Iterations

15. As stated in the Bains Declaration, Plaintiffs have neither first-hand knowledge nor familiarity with the proprietary technology chosen by Defendants nor unfettered interactive access to that technology and the documents sufficient to enable them to gauge the effects of particular relevancy determinations. This is a major disability and disadvantage.

16. From my own work with other predictive coding technology, I have first-hand experience in how interactive testing of different training materials and analysis of the results can make a dramatic difference in the final results. Plaintiffs herein have no such access to the specific technology and data.

17. On the other hand, Defendants have first-hand experience and access to the full document collection as well as privileged access to vendor technical support and expertise.

18. Even as the parties were jointly making relevance determinations, Defendants could run every combination of their relevancy determinations and Plaintiffs' against the entire dataset or random samples. Unlike Plaintiffs, they were in a position to know the effect of each and every relevancy determination made and could plan their strategy accordingly. As Mr.

Baskin stated during the May 14, 2012 Conference, referring to the proprietary system's capabilities, such as phrase extraction, Defendants' "attorneys have all the tools". May 14, 2012 Conference Transcript at 51.

19. For example, what were the specific effects, in this specific data set, of excluding otherwise relevant documents just because they related to a non-plaintiff? Defendants can know this to whatever level of detail they desire while Plaintiffs (and the Court) can only guess.

20. Similarly, only Defendants know what concept groups they used in defining the set of documents to review within a trained category set. May 14, 2012 Conference Transcript at 53.

21. Given Defendants' ultimate production responsibility, enabling Defendants to take full responsibility for the iterative steps, thereby eliminating the seven time-consuming and burdensome joint rounds of negotiations and accompanying court hearings, will yield significant savings in time and cost. *However, Plaintiffs can take this step only if a sound and strong quality control process is assured via inclusion of the validation procedures based upon a random sample of the full collection as set forth above in our proposed protocol.*

A Simple Random Sample From the Entire Collection Is The Keystone of Plaintiffs' Proposal

22. The most critical element of the proposal is having the validating random sample drawn from the full collection rather than the "null set" (Defendants' terminology).

23. As set forth in the Lewis Declaration, Defendants' proposed use of the null set has multiple fatal flaws, e.g.,

1. Defendants' proposed sampling strategy cannot distinguish between effective and ineffective use of predictive coding ("***no approach based on sampling only from the null set can be relied upon to distinguish good from poor predictive coding***", Lewis Declaration ¶ 17 (emphasis in original),

2. Defendants' proposed test for whether a review has achieved high effectiveness is invalid ("**50% of the time in this situation where predictive coding was completely useless, Defendants' approach would indicate that there was some evidence of the success of predictive coding**", Lewis Declaration ¶ 17 (emphasis in original)), and

3. Defendants' proposed sampling strategy has other flaws (including contextual bias, Lewis Declaration ¶ 24) as do other possible variants of that strategy

(including the statistical challenges of validating a stratified sample, Lewis Declaration ¶ ¶ 27-28).

24. Mr. Lewis' objections to the use of elusion based upon the null set to validate a predictive coding process are echoed by others in the eDiscovery industry. Maura R. Grossman and Gordon McCormack were cited in this Court's article *Search Forward* (Andrew Peck, Search, Forward, L. Tech. News, Oct. 2011). They were both coordinators of the TREC Legal Tracks for 2010, 2011 and 2012 (cancelled) and are the authors of the recent commonly cited article *Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review*, in XVII RICH. J.L. & TECH 11 (2011), which was also cited in *Search Forward.* They recently published *The Grossman-Cormack Glossary of Technology Assisted* Review (Version 1.0), October 2012 (available at http://cormack.uwaterloo.ca/targlossary/ and http://www.wlrk.com/webdocs/wlrknew/AttorneyPubs/WLRK.22191.12.pdf). The glossary provides the following definition of elusion:

> "Elusion: The fraction of documents identified as Non-Relevant by a search or review effort, that are in fact Relevant. Elusion is computed by taking a Random Sample from the Null Set and determining how many (or what Proportion of) documents are actually Relevant. *A low Elusion value has commonly been advanced as evidence of an effective search or review effort (see, e.g., Kleen), but that can be misleading because absent an estimate of Prevalence, it conveys no information about the search or review effort. Consider, for example, a document collection containing one million documents, of which ten thousand (or 1%) are Relevant. A search or review effort that found none of the Relevant documents would have 1% Elusion, belying the failure of the search. Elusion = 100% – Negative Predictive Value.*" *Glossary* at 7 (emphasis added).

25. As the Lewis Declaration notes in its discussion of other flaws in Defendants' proposed sampling strategy:

> "Another problem with Defendants' approach to evaluation is that sampling from the null set for a particular prioritization produced by predictive coding only allows evaluation of that single predictive model. *Of even more interest is evaluating the effectiveness of Defendants' entire production process*, including all methods of finding responsive documents, and any losses incurred through the error of reviewers (if Defendants choose to use manual review in their production process). Defendants' approach to evaluation does not allow this latter effectiveness to be estimated." Lewis Declaration ¶ 25 (emphasis added).

26. Plaintiffs' proposal provides for the random sample to be taken from the full collection. Per Lewis:

> "Plaintiffs propose using the standard strategy for evaluating classification systems: take a simple random sample of the entire collection large enough to contain a reasonable number of responsive documents, jointly assess it for responsiveness by assessors who are blinded as to which documents have been prioritized, and estimate effectiveness measures based on this sample. This approach to evaluation is widely used in other applications of text classification outside e-discovery, and more broadly for applications of other types of data classification in industry and science.... A simple random sample of the entire collection can be used to estimate recall, precision, or any other effectiveness measure for binary classification. " Lewis Declaration ¶¶ 31-32.

27. As further set forth in the Lewis Declaration, this approach allows for estimating effectiveness measures proposed by both Defendants and Plaintiffs; as Lewis states as follows:

> **Thus under their own assumptions, Defendants' sampling strategy would not provide a meaningful advantage over Plaintiffs' sampling strategy, even in estimating an effectiveness measure *put forward by Defendants.* What Defendants' proposed strategy *would* do is to make it impossible to estimate recall, i.e. make it impossible to determine how effective predictive coding has been at prioritizing responsive documents for review.** (Emphasis in original.)
> *Measuring Effectiveness and the Fall Off the Cliff*

28. Defendants propose delaying the selection of a quality control sample until after an arbitrary limit on review has been established. Mr. Anders 14-Nov-2012 email states: 'Therefore, MSL proposes that after it completes the iterative rounds of predictive coding, it makes a proposal as to the number of documents to be reviewed for production (e.g., the top ranked 40,000 documents – or whatever number agreed to by the parties or determined by the court). MSL will provide an explanation as to the basis for its proposal. At that point, but prior to final review and production, MSL will generate and the parties will review a second random sample comprised of the remaining documents (e.g., everything other than the top-ranked 40,000 documents – or whatever number agreed to by the parties or determined by the court) to

determine whether it contains any relevant documents and, if so, the nature of those documents (e.g., are they highly relevant or "more of the same.")'.

29. Per Mr. Lewis, this is "truly putting the cart before the horse". Lewis Declaration ¶ 43. Or, in the words of another Lewis, Lewis Carroll, this is "Sentence first – verdict afterward". Particularly where, as here, misplaced reliance on sampling the null set means that the verdict will be useless in any case.

30. Plaintiffs cannot determine if the training strategies proposed by Defendants are designed to or would achieve good effectiveness. Mr. Anders 14-Nov-2012 email states "As it relates to a new seed set, we have been advised that the preparation of a new seed set is not necessary. The current seed set comprises of approximately 16,000 documents which we understand is sufficient for the initial training of the software." We have no knowledge who so advised him, nor any demonstrable and verifiable basis for that advice.

31. However, as Mr. Lewis states:

> "these 16,000 documents were selected from what is now only a subset of a much larger collection. Further, the responsiveness assessments made on those documents are no longer valid. While there was discussion during the 14-May-2012 hearing about how the use of category judgments could be used to compensate for the change in responsive definition from, for instance, the addition of additional plaintiffs, the details of this are not clear to me or to Plaintiffs." Lewis Declaration ¶ 44.

It is not clear to me either.

32. In addition to specifying the simple random sample from the full collection that is the only way to enable calculation of truly validating statistics, Plaintiffs' proposed protocol also ensures that any claims that that relevance fell off a cliff can accurately be assessed.

33. The list specifying the unique ID of every document in the collection and the score or scores assigned to it by the final predictive model or models used by Defendants included in Plaintiffs' protocol will provide additional guidance as to the number of documents to review and, in combination with the quality control random sample, allow statistically validating any claim that the effectiveness of predictive coding has reached any purported "cliff

edge". Like Mr. Lewis, I concur with the Court's statement during the Feb. 8, 2012 hearing: "I'm not saying that there is a dollar number that I'm going to cut it off at or a percentage or where the cliff is. We are going to figure all that out." Plaintiffs' proposal provides accurate means to do just that while Defendants' proposal does not.

Proportionality and Cost

34. Defendants have raised the issues of proportionality and cost in a production larger than 40,000 documents. I note three points in this regard.

35. First, as stated in the Lewis Declaration, Plaintiffs' strategy can be implemented with a reasonably sized sample. Depending on the level of recall attained by Defendants' process and/or the confidence interval selected, this set could be as small as 6,183 documents.

36. Second, a large portion of Defendants' costs appear to stem from manual review of every document in the entire production set. These costs are largely unnecessary and voluntarily assumed by the Defendants. To the extent this type of review is completely unnecessary, the costs associated with any such review should not be considered in determining costs for weighing proportionality or for any other purpose.

37. The entire purpose of advanced techniques such as predictive coding is to reduce manual review. Here the only manual review required is that which is required for training the system and determining relevance in the random samples required to prove the effectiveness of the process.

38. Third, Defendants have rejected an automated privilege review, despite the existence of a Rule 501(d) order.

39. As the Court stated in *10 Key E-Discovery Issues In 2011: Expert Insight to Manage Successfully*:

> "Appropriate use of keywords and/or other automated search techniques can identify most potentially privileged documents – e.g., by searching for words like attorney, counsel, lawyer, legal, privilege; by searching for the names of known in-house and outside counsel; and by searching for email address extensions of outside counsel (e.g.,

any email in the form "name@weil.com"). Such searches are not likely to find an e-mail between two non-lawyer company executives saying something like "David said we should … ," where David is outside counsel, but query whether even a human reviewer would recognize that e-mail as privileged. Moreover, if a Rule 502(d) order has been sought and entered, any such e-mail that falls through the cracks can be recalled without any loss of privilege." David J. Lender & Hon. Andrew J. Peck, *10 Key E-Discovery Issues In 2011: Expert Insight to Manage Successfully,* The Metropolitan Corporate Counsel Northeast Edition, Vol 19, Issue 4, April 2011 at 1, 6.

40. Here we have the precise circumstance set forth in the article. Defendants' decision to reject this course is voluntary and any costs arising there from are thereby ipso facto excessive and should also not be considered in determining costs for weighing proportionality or for any other purpose.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed Nov 15, 2012, at Pelham, NY

Douglas E. Forrest

Exhibit G

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

MONIQUE DA SILVA MOORE, MARYELLEN O'DONOHUE, LAURIE MAYERS, HEATHER PIERCE, KATHERINE WILKINSON on behalf of themselves and all others similarly situated, and ZANETA HUBBARD, on her own behalf.	
PLAINTIFFS,	**Civ No. 11-CV-1279 (ALC) (AJP)**
v.	
PUBLICIS GROUPE SA and MSLGROUP,	
DEFENDANTS.	

DECLARATION OF DEEPIKA BAINS IN SUPPORT OF PLAINTIFFS' PROPOSED CHANGES TO THE ESI PROTOCOL

DEEPIKA BAINS, an attorney duly admitted to practice law in the State of New York, in the Southern District of New York, states as follows:

1. My firm represents the Plaintiffs in the above-referenced action.

2. Since the Court's adoption of the current ESI protocol, Plaintiffs' counsel have diligently worked with MSL's counsel to ensure its smooth implementation. However, the predictive coding implementation process has been much more time-consuming and burdensome than traditional discovery methods for the parties and the Court.

3. Neither Plaintiffs nor their counsel has first-hand knowledge or familiarity with MSL's proprietary technology used to implement the predictive coding process.

4. With respect to only the first step of a nine step process (i.e., the *initial* random sample and the seed set), the parties expended an enormous amount of time and resources not only reviewing over 15,000 documents and coding those documents, but also meeting and conferring regarding the coding discrepancies. Some of the time-consuming steps taken, among others, include:

a. MSL produced its initial random sample on March 21, 2012. MSL then produced several batches of seed set documents on March 28, March 30, April 3, April 6 and April 11.
b. Plaintiffs reviewed the initial random sample and seed set documents and submitted their coding changes to MSL on April 11, 2012 and April 23, 2012.
c. After making initial coding designations, Plaintiffs were required to re-review such designations again after the April 24, 2012 court conference.
d. Within 24-hours of the conference, Plaintiffs were required to send a list of inconsistently coded documents (documents coded as both relevant and non-relevant in different productions) on April 25, 2012.
e. On April 30, 2012, Plaintiffs submitted their revised coding changes to MSL.
f. During one particular one-week period (May 4, 2012 through May 10, 2012), the parties held three lengthy meet and confers (some lasting several hours) in an attempt to resolve coding disputes. During this time, Plaintiffs and Defendant exchanged numerous emails regarding coding designations.
g. MSL sent a response to Plaintiffs' coding designations on May 3, 2011 and responded to Plaintiffs' list of inconsistently coded documents on May 6, 2012. That day, Plaintiffs reviewed and responded to MSL's coding designations with their own changes. Plaintiffs submitted further changes on May 9, 2012.

5. The Court has also expended an unusual amount of time on this matter because predictive coding is being used. For example, following the parties' joint review of the initial random sample and seed set documents, the Court devoted three conferences in as many weeks (April 25, 2012, May 7, 2012 and May 14, 2012) to resolving coding disputes and other issues relating to the implementation of the protocol. At the time the Court issued a stay of the ESI protocol, on May 14, 2012, there still remained several hundred documents in dispute from the seed set alone.

6. Plaintiffs anticipate that the current protocol's seven iterative rounds and final random sample will also give rise to numerous coding disputes, requiring several meet and confers and Court conferences.

7. Having experienced firsthand the extensive time and burden associated with implementing simply the *initial* phase of the protocol, and recognizing that the addition of 26 new opt-in Plaintiffs will necessarily broaden the scope of discovery and the protocol at this stage (and class certification or additional plaintiffs will again broaden the scope at a later stage), Plaintiffs continue to maintain that traditional discovery methods are more appropriate for this employment discrimination class and collective action.

8. Nonetheless, Plaintiffs have consulted their experts with a view to streamlining the predictive coding process adopted by the Court. Plaintiffs' proposed changes, which are detailed in Plaintiffs' letter and in the Declarations of David D. Lewis, Ph.D., and Douglas Forrest, will result in a more effective predictive coding review while minimizing the burden on the parties and the Court.

9. Notably, Plaintiffs' proposal reduces the number of joint steps in the protocol from nine (initial random sample/seed set; seven iterative rounds; final random sample) to two (initial random sample/seed set; final random sample). This reduction in Plaintiffs' review is contingent on Plaintiffs' ability to do a thorough and complete review of the final random sample, as described in detail in Plaintiffs' letter and the Lewis and Forrest Declarations. Although Plaintiffs' proposal does not require joint review of the seven iterations of documents, Plaintiffs will make themselves available for any questions from or discussions with MSL counsel, as needed.

10. Reducing the number of joint steps in the protocol and shortening the schedule not only promotes judicial efficiency, but also serves Plaintiffs' interest in promptly receiving responsive documents. So far, Plaintiffs have been unable to take depositions or otherwise advance the case because they have not yet received the vast majority of e-discovery from MSL.

Plaintiffs' Proposed Changes to Scope of Protocol in Light of 26 New Opt-in Plaintiffs

11. At the time the Court adopted the ESI protocol, there were seven plaintiffs in the case: Named Plaintiffs Monique da Silva Moore, MaryEllen O'Donohue, Laurie Mayers, Heather Pierce, and Kate Wilkinson, and opt-in Plaintiffs Carol Perlman and Zaneta Hubbard (Ms. Hubbard later became a Named Plaintiff).

12. With some exceptions (e.g., policy documents), the Court has limited discovery in the case to email documents relating to the abovementioned Plaintiffs.

13. On May 14, 2012, Judge Peck stayed e-discovery pending a decision on Plaintiffs' conditional certification motion, "in light of the fact that if additional plaintiffs opt in, that will change what documents are responsive." May 14, 2012 Tr. at 74.

14. On June 29, 2012, the Court granted conditional certification of the Equal Pay Act (EPA) collective action. Twenty-six additional women have since opted into the collective action and are now EPA opt-in Plaintiffs.

15. In his October 11, 2012 Order, Judge Peck ordered the parties to "discuss any appropriate changes to the ediscovery protocol or procedures related to the opt-ins."

16. Consistent with the Court's Order, the parties held a telephonic meet and confer regarding their proposed changes on November 12, 2012. Plaintiffs' Counsel also exchanged e-mails with defense counsel regarding the parties' proposed changes to the protocol on November 12, 13 and 14, 2012. Although the parties appeared to be in agreement that the addition of 26 new opt-ins necessitated some changes to the ESI protocol, we were unable to agree on certain fundamental issues. Accordingly, the parties agreed to submit separate proposals to the Court.

17. Plaintiffs believe the addition of 26 new opt-in Plaintiffs changes the scope of the protocol in the following significant ways:

    a. **Collection of Additional Accounts:** The e-mail accounts of the 26 new opt-in Plaintiffs and their comparators will need to be added to the collection. The opt-in group grew by approximately 500% as compared to the group before the notice period.

    b. **Temporal Scope:** Because the opt-in group has expanded significantly and because there are at least three current employee opt-in plaintiffs, the discovery cutoff should extend past MSL's proposal of November 1, 2012. The class period has not yet been set, and MSL is under a continuing obligation to produce relevant discovery for all class members. The temporal scope must be expanded to accommodate newly added opt-in Plaintiffs and future plaintiffs to be added after the Court's decision on class certification.

    c. **Scope of Relevance:** Responsiveness must be expanded to include all documents related to the pay, promotion, and pregnancy claims of all current opt-in Plaintiffs. Additionally, rather than re-doing the ESI protocol once again after a decision is made regarding class certification, Plaintiffs maintain that it is more efficient to adopt a definition of responsiveness that incorporates Title VII class members.

I DECLARE UNDER THE PENALTIES OF PERJURY THAT THE FOREGOING IS TRUE AND ACCURATE TO THE BEST OF MY KNOWLEDGE AND BELIEF.

Dated: November 15, 2012

New York, New York

/s/ Deepika Bains

Deepika Bains